



中华人民共和国国家标准

GB 4883—85

数据的统计处理和解释 正态样本异常值的判断和处理

Statistical interpretation of data—Detection and
handling of outlying observations in normal sample

1985-01-29 发布

1985-10-01 实施

国家标准局 批准

数据的统计处理和解释

正态样本异常值的判断和处理

Statistical interpretation of data—Detection and handling of outlying observations in normal sample

1 引言

1.1 本标准规定了判断和处理在正态样本中出现的异常值的一般原则和实施方式。

1.2 异常值(或异常观测值)是指样本中的个别值,其数值明显偏离它(或它们)所属样本的其余观测值。

异常值可能是总体固有的随机变异性的极端表现。这种异常值和样本中其余观测值属于同一总体。

异常值也可能是由于试验条件和试验方法的偶然偏离所产生的后果,或产生于观测、计算、记录中的失误。这种异常值和样本中其余观测值不属于同一总体。

1.3 本标准使用的其它统计学名词,见国家标准GB 3358—82《统计学名词及符号》。

1.4 应用条件:所考查样本中诸观测值(或经过一定的函数变换后得到的值),除了个别异常值外,其余大部分值(样本主体)来自同一正态总体或近似正态总体。

关于样本来自正态总体或近似正态总体的判断,可以根据物理上的、技术上的知识;也可通过与考查对象有同样性质的以往数据,进行正态性检验,其原理和方法见国家标准GB 4882—85《数据的统计处理和解释—正态性检验》。

2 判断异常值的统计学原则

2.1 本标准在下述不同情形下判断样本中的异常值:

上侧情形:根据以往经验,异常值都为高端值;

下侧情形:根据以往经验,异常值都为低端值;

双侧情形:异常值是在两端都可能出现的极端值。

注:上侧情形和下侧情形统称单侧情形。

2.2 执行本标准时,应规定在样本中检出异常值的个数的上限(占样本观测值个数的较小比例),当超过了这个上限,对此样本的代表性,应作慎重的研究和处理。

2.3 判断单个异常值的检验规则

根据实际情况,选定适宜的异常值检验规则(见4、5、6章);

指定为检出异常值的统计检验的显著性水平 α ,简称检出水平;

根据 α 和观测值个数 n 确定统计量的临界值;

将各观测值代入检验规则中给出的统计量,所得值若超过临界值,则判断事先确定待查的极端观测值为异常值;否则就判断“没有异常值”。

检出水平 α 的宜取值是5%,1%(或10%)。

2.4 判断多个异常值的检验规则

在允许检出异常值个数可大于1的情形,本标准规定的方法是重复使用同一种判断单个异常值的检验规则,即用指定的检出水平和符合2.3规定的规则首先检验全体观测值,若不能检出异常值,则整个检验停止;若检出了一个异常值,就再用相同的检出水平和相同的规则,对除去已检出的异常值后余下的观测值继续检验……直到不能检出异常值,或检出的异常值个数超过上限为止。